

Power Allocation with Stackelberg Game in Femtocell Networks: A Self-Learning Approach

Wenbo Wang*, Andres Kwasinski* and Zhu Han†

*Department of Computer Engineering, Rochester Institute of Technology, NY, USA

†Department of Electrical and Computer Engineering, University of Houston, TX, USA

Abstract—This paper investigates the energy-efficient power allocation for a two-tier, underlaid femtocell network. The behaviors of the Macrocell Base Station (MBS) and the Femtocell Users (FUs) are modeled hierarchically as a Stackelberg game. The MBS guarantees its own QoS requirement by charging the FUs individually according to the cross-tier interference, and the FUs respond by controlling the local transmit power non-cooperatively. Due to the limit of information exchange in intra- and inter-tiers, a self-learning based strategy-updating mechanism is proposed for each user to learn the equilibrium strategies. In the same Stackelberg-game framework, two different scenarios based on the continuous and discrete power profiles for the FUs are studied, respectively. The self-learning schemes in the two scenarios are designed based on the local best response. By studying the properties of the proposed game in the two situations, the convergence property of the learning schemes is provided. The simulation results are provided to support the theoretical finding in different situations of the proposed game, and the efficiency of the learning schemes is validated.

I. INTRODUCTION

Femtocells are considered to be an efficient and economic solution to enhance the indoor experience of the cellular mobile users [1]. A femtocell is a low-power, short-range access point, which can be quickly deployed by the end-users. It provides better spatial reuse of the spectrum by serving the nearby users who have poorer connections with the Macrocell Base Station (MBS) due to penetration loss. In practice, the femtocell network usually operates underlying the macrocell network. This is mainly due to the ad-hoc topology of the femtocells and thus the lack of coordination between the MBS and Femto Access Points (FAPs). Consequently, inter-cell/cross-tier interference arises, and interference mitigation becomes necessary for preventing performance deterioration.

Due to the ad-hoc topology of the femtocells, the FAP deployment faces the limited information exchange both across tiers and among the femtocells. Therefore, it is desirable that the interference management of the femtocells is fully distributed, and each Femtocell User (FU) is capable of adapting to the surrounding environment with minimum information. With this in mind, we study the power control schemes for a shared-spectrum, two-tier femtocell network. We note that the Macrocell User (MU) prefers that the cross-tier interference is minimized, while the FUs prefer to transmit with the best Signal-to-Interference-plus-Noise-Ratio (SINR). Considering that private objectives contradict with each other, it is natural to introduce the tools of game theory and model the cross-tier, self-centric interactions in the framework of non-cooperative games.

A. Related Work

Under the framework of non-cooperative games, the early study [2] has discovered that power control purely based on the non-cooperative games will lead to inefficient equilibria. In order to obtain the Pareto-preferred equilibria, a number of approaches including the introduction of repetition [3] or externalities (e.g., pricing) [4], [5] are adopted in the research. As shown by studies in non-hierarchical networks [4]–[6], choosing a proper pricing mechanisms with respect to different utility functions can be an efficient way of determining the desired properties of the equilibria.

When it comes to the resource allocation in hierarchical networks, such as the femtocell networks and cognitive radio networks, the Stackelberg game [7] based modeling is widely preferred since it is able to reflect the features of hierarchy and ad-hoc topology in the network [8], [9]. The Stackelberg game is characterized by the sequential decision making manner (namely, the follower-leader strategy updating), and hence suitable for modeling the heterogeneous user behaviors in the network. Due to the computational intensity for obtaining the Stackelberg Equilibrium (SE), most of the existing studies [9]–[11] adopt a utility model that favors the derivation of a closed-form SE, and solve for the SEs through transforming the games as hierarchical optimization problems.

Although the optimization-technique-based methods are able to precisely analyze the properties of the SEs, their scope is limited to the games with a certain categories of utility functions. Beyond those games, a natural idea is to resort to tools of iterative learning in repeated games for searching the SEs. A body of literature on non-hierarchical networks can be found applying iterative strategy-learning methods [12], [13], usually based on the assumption of the discrete strategy space [14]. However, since these learning methods assume homogeneous behaviors among the players, most of their application to hierarchical networks are also limited within an uniform learning model [15], [16].

B. Self-Learning under Pricing

In this paper, we model the power allocation problem in the two-tier femtocell network from the perspective of the Stackelberg game. In the game, the MBS behaves as the leader and controls the total cross-tier interference by setting prices to each FU-FAP link. The FU-FAP links behave as the followers to optimize their energy efficiency through interactive power allocation. In designing the distributed, hierarchical power control scheme with pricing, the power efficiency is adopted

as the utility of the FU-FAP link. we investigate the scenarios of the continuous and discrete power profile of the femtocells, respectively. Due to the computational complexity of obtaining the closed-form solution of the equilibrium, we investigate the property of the game in the two situations, and propose two self-centric strategy-learning algorithms for the follower game based on the local myopic best response, respectively. We provide the theoretical proof of the convergence of the learning schemes. Also, we provide two heuristic algorithms for obtaining the optimal MBS prices in the corresponding situations. In the simulation, we provide the experimental comparison on the network performance between the two scenarios, and demonstrate the efficiency of the proposed learning algorithms.

II. PROBLEM FORMULATION

A. Network Model

We consider the uplink transmission of a two-tier femtocell network with a single MBS and K FAPs. The MBS and the FAPs share the same bandwidth W and each of them is scheduled to serve one single user at each time instance. The MBS is required to keep the femtocell-to-macrocell interference to an acceptable level. Due to the ad-hoc topology of the femtocells, we assume that the information exchange only happens between the MBS and FAPs. For analytical tractability, we suppose that all the channels involved are block-fading and remains constant during each transmission block.

In what follows, we let 0 denote the index of the MU-MBS pair and $k \in \mathcal{K} = \{1, \dots, K\}$ denote the index of a FU-FAP pair. The channel power gain between the transmitter of pair i and the receiver of pair j is denoted by $h_{i,j}$, where $i, j \in \mathcal{K} \cup \{0\}$. The power of transmitter k is denoted by p_k and the power vector of all the FUs is denoted by $\mathbf{p} = [p_1, \dots, p_K]$. The noise variance for the transmitter-receiver pair k is denoted by N_k . Then the SINR level at the MBS can be expressed as:

$$\gamma_0(p_0, \mathbf{p}) = \frac{h_{0,0}p_0}{N_0 + \sum_{k \in \mathcal{K}} h_{k,0}p_k}, \quad (1)$$

and the SINR level at the k -th FAP can be expressed as

$$\gamma_k(p_0, \mathbf{p}) = \frac{h_{k,k}p_k}{N_k + h_{0,k}p_0 + \sum_{j \in \mathcal{K} \setminus \{k\}} h_{j,k}p_j}. \quad (2)$$

During the operation, it is usually beneficial to shift some calls served by the MBS to the FAP. Therefore, we suppose that for the MBS, the requirement on the femtocell-to-macrocell interference is not rigid. Instead, the MU transmit with a fixed power and the MBS charges each FU-FAP link for causing interference with a certain price to control the interference level. We denote the vector of interference prices at a time interval by $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_K]$, in which λ_k is the price for unit interference caused by FU k . The goal of the MBS is to maximize the total revenue of collecting payments from the FU-FAP links:

$$\max_{\boldsymbol{\lambda}} \left(u_0 = \sum_{k \in \mathcal{K}} \lambda_k h_{k,0} p_k \right). \quad (3)$$

For simplicity, in what follows we use the terms FU and the FU-FAP link interchangeably. For the FUs, we assume that each local transmit power p_k is limited by the physical constraint $p_k \in [0, p_k^{\max}]$. The goal of FU k is to maximize its local net payoff by adapting p_k :

$$\max_{0 \leq p_k \leq p_k^{\max}} \left(u_k = \psi(\gamma_k, p_k) - \lambda_k h_{k,0} p_k \right), \quad (4)$$

in which $\psi(\gamma_k, p_k)$ is the utility function of FU k . Considering the practical scenarios, we adopt the local utility $\psi(\gamma_k, p_k)$ as the energy efficiency (namely, the data received per unit energy consumption) of each FU:

$$\psi(\gamma_k, \mathbf{p}_k) = \frac{W \log(1 + \gamma_k)}{p_k + p_a}, \quad (5)$$

where p_a denotes the additional circuit power consumption for FU k . We assume that the local SINR γ_k can be perfectly measured at the FAP. For the proposed femtocell network, we suppose that the following assumptions hold:

- i) p_k is significantly greater than p_a ;
- ii) The femtocell-to-femtocell interference is sufficiently small.

Assumption i) is based on the fact that most power is consumed during operation by the amplifier and the radio transceiver [17]. Assumption ii) is based on the practical concerns that (a) the FAPs are of low power, so the peak transmit power is limited; and (b) the intercell gains between femtocells are usually weak due to the path loss (since the indoor penetration loss is usually significant). Based on the assumptions, we assume that the SINR constraint for a FU-FAP link is negligible.

III. STACKELBERG GAME ANALYSIS

We model the user interactions in the proposed network as a hierarchical game with the MBS and the FU-FAP links choosing their actions in a sequential manner. When the power allocation of the FUs are given as \mathbf{p} , we can define the leader game from the perspective of the MBS as:

$$\mathcal{G}_l = \langle \boldsymbol{\Lambda}, \{u_0(\boldsymbol{\lambda}, \mathbf{p})\} \rangle, \quad (6)$$

in which the MBS is the only player of the game and the action of the player is the price vector $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$.

When the leader action is given by $\boldsymbol{\lambda}$, we can define the non-cooperative follower game from the perspective of the FUs as:

$$\mathcal{G}_f = \langle \mathcal{K}, \mathcal{P}, \{u_k(\boldsymbol{\lambda}, \mathbf{p}_k, \mathbf{p}_{-k})\}_{k \in \mathcal{K}} \rangle, \quad (7)$$

in which FU k is one player with the local action $\mathbf{p}_k \in \mathcal{P}_k$.

With each player behaving rationally, the goal of the game (6) and (7) is to finally reach the Stackelberg Equilibrium (SE) in which both the leader (MBS) and the followers (FUs) have no incentive to deviate. We first assume that the strategies of the FUs and MBS are continuous. Then the SE of the game can be mathematically defined as follows:

Definition 1 (Stackelberg Equilibrium). *The strategy $(\boldsymbol{\lambda}^*, \mathbf{p}^*)$ is a SE for the proposed Stackelberg game (6) and (7) if*

$$u_0(\boldsymbol{\lambda}^*, \mathbf{p}^*) \geq u_0(\boldsymbol{\lambda}, \mathbf{p}^*), \forall \boldsymbol{\lambda} \in \boldsymbol{\Lambda}, \quad (8)$$

$$u_k(\boldsymbol{\lambda}^*, \mathbf{p}_k^*, \mathbf{p}_{-k}^*) \geq u_k(\boldsymbol{\lambda}^*, \mathbf{p}_k, \mathbf{p}_{-k}^*), \forall k \in \mathcal{K}, \forall \mathbf{p}_k \in \mathcal{P}_k. \quad (9)$$

A. Femtocell Power Allocation with Continuous Strategies

We start the analysis of the Stackelberg game in (6) and (7) by back induction. Suppose that the MBS first sets its strategy as λ , then we obtain the non-cooperative follower subgame described by (7). In order to show the existence of a pure-strategy Nash Equilibrium (NE) in \mathcal{G}_f , we introduce the concept of supermodular game as follows.

Definition 2 (Supermodularity [7]). *A function $f : \mathcal{X} \times \mathcal{T} \rightarrow \mathbb{R}$ is said to have increasing differences (supermodularity) in (x, t) if for all $x' \geq x$ and $t' \geq t$,*

$$f(x', t') - f(x, t') \geq f(x', t) - f(x, t).$$

Definition 3 (Supermodular game [7], [18]). *A general normal-form game $\langle \mathcal{N}, \{\mathcal{S}_i\}_{i \in \mathcal{N}}, \{u_i\}_{i \in \mathcal{N}} \rangle$ is a supermodular game if for any player $i \in \mathcal{N}$,*

- i) *the strategy space \mathcal{S}_i is a compact subset of \mathbb{R}^K .*
- ii) *the payoff function u_i is upper semi-continuous in $\mathbf{s} = (\mathbf{s}_i, \mathbf{s}_{-i})$.*
- iii) *u_i is supermodular in \mathbf{s}_i and has increasing difference between any component of \mathbf{s}_i and any component of \mathbf{s}_{-i} .*

The supermodular property of the proposed follower subgame (7) is given by Theorem 1.

Theorem 1. *Given the strategy λ of the MBS, the follower subgame (7) is a supermodular game if $\gamma_k \geq p_a/p_k$.*

Proof. See Appendix A. \square

Given the opponent strategies p_{-k} , we define the local Best Response (BR) of FU k by

$$\hat{p}_k(p_{-k}) = \arg \max_{0 \leq p_k \leq p_k^{\max}} \left(u_k = \psi(\gamma_k, p_k) - \lambda_s h_{k,0} p_k \right). \quad (10)$$

Then Theorem 1 immediately yields Proposition 1 [18]:

Proposition 1. *At least one pure-strategy NE exists in the follower subgame (7) and the following points hold:*

- i) *The set of NEs (9) has the component-wise greatest element $\bar{\mathbf{p}}^*$ and least element $\underline{\mathbf{p}}^*$.*
- ii) *If the BRs are single-valued, and each FU uses the BR starting from the smallest (largest) elements of the strategy space to update their strategies, then the strategies monotonically converge to the smallest (largest) NE.*
- iii) *If the game has a unique NE, then with any arbitrary initial strategy, the local myopic BRs converges to the NE.*

The properties of \mathcal{G}_f in Proposition 1 sheds light on the solution to the local strategy-learning scheme for the FUs. To take advantages of these properties, we first examine the BRs in \mathcal{G}_f and obtain Lemma 1:

Lemma 1. *Given the MBS strategy λ , u_k is strictly quasiconcave and the best-response \hat{p}_k is single-valued for each FU.*

Proof. See Appendix B. \square

Based on Lemma 1, we can further prove that the BR of FU k has the following properties, and therefore is a standard function [7]:

- i) positivity: $\hat{p}_k > 0$,
- ii) monotonicity: for any $p'_{-k}, p_{-k} \in \mathcal{P}_{-k}$, if $p'_{-k} > p_{-k}$, then $\hat{p}_k(p'_{-k}) \geq \hat{p}_k(p_{-k})$,
- iii) scalability: $\forall \alpha > 1$, $\alpha \hat{p}_k(p_{-k}) > \hat{p}_k(\alpha p_{-k})$.

Based on the aforementioned properties, we can directly apply the results in [7] and deduce the uniqueness of the pure-strategy NE in the follower subgame \mathcal{G}_f .

Theorem 2. *Given any MBS strategy λ , the follower subgame (7) has a unique NE if the following condition is satisfied*

$$\frac{h_{k,k} p_k}{N_k + h_{0,k} p_0 + \sum_{j \in \mathcal{K}} h_{j,k} p_j} \geq \frac{p_a}{p_k}. \quad (11)$$

Proof. See Appendix C. \square

Remark 1. If in subgame \mathcal{G}_f the condition of Theorem 2 is satisfied, the condition of Theorem 1 will also be satisfied. (11) indicates that to ensure the uniqueness of the NE, the SINR of a FU should be significantly larger than the ratio between its transmit power and circuit power. Such a condition is guaranteed by our assumption of the network.

We assume that the channel power gain $h_{k,0}$ is known and the SINR γ_k can be perfectly measured by each FAP. Then, based on Lemma 1, \hat{p}_k can be solved locally with the bisection method [19]. Based on Proposition 1 and Theorem 2, the asynchronous strategy-updating mechanism defined in [18] can be directly applied to \mathcal{G}_f . By Proposition 1, the convergence to the NE is guaranteed from any arbitrary initial power vector. The strategy-learning algorithm is summarized as Algorithm 1:

Algorithm 1 Asynchronous strategy updating

Require: each FU sets up an infinite increasing time sequence

- $\{T_k^i\}_{k \in \mathcal{K}}$ for scheduling strategy update.
 - 1: **for all** $t \in \{T_k^i\}_{k \in \mathcal{K}}$ **do**
 - 2: **for all** k s.t. $t = \{T_k^i\}$ **do**
 - 3: given p_{-k}^{t-1} , obtain $p_k^t = \hat{p}_k(p_{-k}^{t-1})$ as in (10) with bisection.
 - 4: **end for**
 - 5: **end for**
-

B. Approximate Solution to the Price of MBS

When considering the leader subgame \mathcal{G}_l , we assume that the strategies of the FUs are given as \mathbf{p} from (10). For the subgame (6), the local BR is given by

$$\hat{\lambda} = \arg \max_{\lambda \geq 0} \left(\sum_{k \in \mathcal{K}} \lambda_k h_{k,0} p_k \right). \quad (12)$$

To investigate the solution of (12), we first consider the feasible region of λ_k . We note from (10) that the maximum value of u_k is lower-bounded by 0 (when $p_k = 0$) and upper-bounded by $\psi(\tilde{\gamma}_k, \tilde{p}_k)$, in which $(\tilde{\gamma}_k, \tilde{p}_k) = \arg \max \psi(\gamma_k, p_k)$. Thereby, the price λ_k charged by the MBS is also upper-bounded.

Otherwise, if λ_k is too high (i.e., making $\psi(\tilde{\gamma}_k, \tilde{p}_k) < \lambda_s h_k p_k$), FU k will stop transmitting and be forced out of the game. With the aforementioned bound on u_k , we look for the constraint on λ_k in (10). However, since u_k in (10) is a transcendental function, it is difficult to derive a closed-form expression of the constraint on λ_k . Then, the challenge of analyzing \mathcal{G}_l is to find an efficient way for obtaining the optimal price $\tilde{\lambda}$.

By jointly investigating the leader and the follower subgames, we can show that a finite, optimal price $\hat{\lambda}_k$ for each FU coexists with the NE of the follower subgame.

Theorem 3. *In the Stackelberg game defined by (6) and (7), at least one pure-strategy SE with finite price vector $\hat{\lambda}$ from the leader game exists.*

Proof. The proof is based on Theorem 3.2 of [7]. Given the condition that each local strategy in the game is compact and convex and the corresponding payoff function is quasiconcave, the existence of a pure-strategy SE is guaranteed. For the FUs, quasiconcavity of $u_k(\mathbf{p}, \boldsymbol{\lambda})$ in p_i is given by Lemma 1. For the MBS, u_0 is an affine function of $\boldsymbol{\lambda}$, hence being quasiconcave in $\boldsymbol{\lambda}$. It is trivial that \mathbf{p} and $\boldsymbol{\lambda}$ are convex and compact, then based on Theorem 3.2 of [7], there exists a pure-strategy NE in the game. Beyond the discussion after (12) on the fact that $0 \leq \lambda_k < \infty$, we can derive the relationship of the BRs between the FUs and the MBS from $\frac{\partial u_k}{\partial p_k} = 0$ as:

$$\tilde{\lambda}_k = \frac{W \tilde{G}_k}{h_k(1 + \tilde{\gamma}_k)(\tilde{p}_k + p_a)} - \frac{W \log(1 + \tilde{\gamma}_k)}{h_k(\tilde{p}_k + p_a)^2}, \quad (13)$$

in which \tilde{p}_k and $\tilde{\gamma}_k$ are the local BR and the corresponding SINR. \tilde{G}_k is given by (27) in Appendix A. For (13), the value of the right-hand side expression is upper-bounded since $0 \leq \tilde{p}_k \leq p_k^{\max}$. Then $\tilde{\lambda}_k$ is finite. \square

Our approximate solution to the leader subgame is inspired by the pioneering work of [4], which models the asymptotic behaviors of the equilibrium bhum power vector and the corresponding payments. We assume that each FU's behavior can be asymptotically modeled by two regions, the price-insensitive region and the price-sensitive region. In the price-insensitive region, the FU's behaviors are hardly influenced by the price. In the price-sensitive region, the local power allocation p_k is driven toward 0. Mathematically, the two-region model can be expressed by the following asymptotes:

- Low-price asymptote as $\lambda_k \rightarrow 0$:

$$\begin{cases} \gamma_k^* \approx \tilde{\gamma}_k, \\ r_k(\lambda_k, p_k) = h_{k,0} \tilde{p}_k \lambda_k \propto \lambda_k. \end{cases} \quad (14)$$

- High-price asymptote as $\lambda_k \rightarrow \infty$:

$$p_k \approx \frac{W}{\lambda_k(N_k + h_{0,k}p_0)} - p_a. \quad (15)$$

In (14), $\tilde{\gamma}_k$ is the equilibrium SINR when $\lambda_k = 0$. The details for deriving (14) and (15) is presented in Appendix D. Since in the low-price asymptote, r_k increases with λ_k and in the high-price asymptote it decreases with λ_k , the maximum payment must happen between the two regions. Then, we can extend

the two FU payment asymptotes toward each other until they meet at the intersection price λ_k^a . With such an approximation, $r_k(\lambda_k^a, p_k)$ will be the maximum payment received from FU k . Combining (14) and (15), we can obtain the intersection point for the two asymptotes as:

$$\lambda_k^a \approx \frac{W}{(p_k^* + p_a)(N_k + h_{0,k}p_0)}, \quad (16)$$

in which p_k^* is the power allocation corresponding to the SINR γ^* in (14). It can be obtained by setting $\lambda = 0$ and solving (40) with the BR-based asynchronous strategy-updating mechanism.

C. Femtocell Power Allocation in Discrete Strategies

We continue to extend the analysis of the game to the scenario in which the FUs choose their strategies from a finite, discrete set of powers. In this case, the conditions for NEs (i.e., Theorems 2 and 3) are not satisfied anymore. Therefore, the properties of the NE need to be re-evaluated. Within the same game structure of (6) and (7), we denote the action set of the FUs by $\mathcal{P}_k = \{p_k^1 = 0, \dots, p_k^{|\mathcal{P}_k|}\}$. It is well known that every finite non-cooperative game has a mixed-strategy NE [7]. For the follower subgame, we define the mixed-strategies of FU k as $\boldsymbol{\pi}_k = [\pi_k^1, \dots, \pi_k^{|\mathcal{P}_k|}]$, in which $\pi_k^j(p_k^j) = \Pr(p_k = p_k^j)$ is the probability for FU k to choose the j -th action $p_k^j \in \mathcal{P}_k$. Then, given any MBS price $\boldsymbol{\lambda}$, there will be at least one mixed-strategy NE for the FUs. Different from the continuous game, the expected net payoff of FU k becomes

$$u_k(\boldsymbol{\pi}_k, \boldsymbol{\pi}_{-k}, \boldsymbol{\lambda}) = \sum_{\mathbf{p} \in \mathcal{P}} (\psi_k(p_k, \mathbf{p}_{-k}) - \lambda_k h_{k,0} p_k) \prod_{i \in \mathcal{K}} \prod_{1 \leq j \leq |\mathcal{P}_i|} \pi_i^j, \quad (17)$$

in which $\sum_j \pi_k^j = 1, 0 \leq \pi_k^j \leq 1$. Similarly, the expected revenue of the MBS becomes

$$u_0(\boldsymbol{\lambda}, \boldsymbol{\pi}) = \sum_{k \in \mathcal{K}} \lambda_k \sum_{1 \leq j \leq |\mathcal{P}_k|} \pi_k^j h_{k,0} p_k^j. \quad (18)$$

Due to the limit of information exchange, FU k can only attain its local payoff $u_k(p_k^j, \boldsymbol{\pi}_{-k})$ each time when it chooses p_k^j and the other FUs adopt the mixed strategies $\boldsymbol{\pi}_{-k}$. To ensure that each FU is able to learn its Nash distribution $\boldsymbol{\pi}_k$, we adopt the Logit best response function (namely, the smooth BR based on entropy perturbation) [20]:

$$\beta_k^t(p_k^j | \boldsymbol{\pi}_{-i}) = \frac{\exp(U_k^{t-1}(p_k^j, \boldsymbol{\pi}_{-k})/\tau)}{\sum_{1 \leq i \leq |\mathcal{P}_k|} \exp(U_k^{t-1}(p_k^i, \boldsymbol{\pi}_{-k})/\tau)}, \quad (19)$$

in which U_k^t is the estimated expected payoff at time t . τ is a positive scalar (also known as Boltzmann temperature) that controls the sensitivity of the BR to perturbation.

Based on the two-timescale strategy learning scheme [20], we introduce two coupled stochastic learning processes to approximate $U_k(p_k^j)$ and $\pi_k(p_k^j)$ in (19) as follows:

$$U_k^t(p_k^j) = U_k^{t-1}(p_k^j) + \alpha_1^t \mathbb{1}_{(\pi_k^j(p_k^j))} \left(u_k^t(p_k^j) - U_k^{t-1}(p_k^j) \right), \quad (20)$$

$$\pi_k^t(p_k^j) = \pi_k^{t-1}(p_k^j) + \alpha_2^t \left(\beta_k^t(p_k^j | \boldsymbol{\pi}_{-i}) - \pi_k^{t-1}(p_k^j) \right). \quad (21)$$

In (20) and (21), $u_k^t(p_k^j)$ is the instant payoff observation at the FAP in (4) and $\beta_k^t(p_k^j|\pi_{-i})$ is the smooth BR (19). $\mathbb{1}_{(\pi_k^j(p_k^j))}$ is the indicator function. $\mathbb{1}_{(\pi_k^j(p_k^j))} = 1$ if $\pi_k^j(p_k^j) = 1$ and otherwise $\mathbb{1}_{(\pi_k^j(p_k^j))} = 0$. The parameter sequence α_1^t and α_2^t satisfy the following conditions:

$$\begin{cases} \lim_{T \rightarrow 0} \sum_{t=1}^T \alpha_1^t = +\infty, & \lim_{T \rightarrow 0} \sum_{t=1}^T (\alpha_1^t)^2 < +\infty, \\ \lim_{T \rightarrow 0} \sum_{t=1}^T \alpha_2^t = +\infty, & \lim_{T \rightarrow 0} \sum_{t=1}^T (\alpha_2^t)^2 < +\infty, \\ \lim_{t \rightarrow 0} \frac{\alpha_t}{\alpha_1} = 0. \end{cases} \quad (22)$$

The conditions in (22) ensures that the learning of strategies changes on a slower timescale than that of the action values. Based on the discussion in [21], we can show that the learning processes converges by Theorem 4:

Theorem 4. *With any arbitrary π^0 and U_k^0 , the strategy-learning mechanism (19)-(21) almost surely converges to some fixed point. The probability of converging to a NE is non-zero.*

Proof. See Appendix E. \square

In the scenario of finite strategies, it is even more difficult to obtain a SE point for the MBS prices. However, by investigating the property of concavity in the payoff function for any element π_k^j or λ_k of the joint strategy vector (π, λ) , we can show that there exists at least one SE with MBS price in pure-strategy:

Theorem 5. *With the discrete set of \mathcal{P}_k , the Stackelberg game with the payoff functions (17) and (18) has a SE composed of the mixed-strategy power allocation $\hat{\pi}$ and the pure-strategy price $\hat{\lambda}$, which is finite in each $\hat{\lambda}_k$.*

Proof. The proof for the existence of the SE follows directly from Theorem 1 of [22]. It is easy to see that the payoff function $U_k(\pi, \lambda)$ is linear (hence concave) in π_k^i if λ and the rest of the elements of π are fixed. Similarly, $U_0(\pi, \lambda)$ is linear in λ_k with λ_{-k} and π fixed. Therefore, following the discussion of Theorem 1 in [22] and the Kakutani fixed point theorem, there exists a fixed point (π^*, λ^*) satisfying Definition 1.

The proof of a finite λ_k in the SE is similar to that in Theorem 3. If we assume that (λ, π) is a SE, then from (17) we obtain $\forall k \in \mathcal{K}, 1 \leq j \leq |\mathcal{P}_k|, \frac{\partial u_k}{\partial \pi_k^j} = 0$, which is equivalent to

$$\sum_{p-k} \psi(p_k^j, p-k) \prod_{m \neq k, i} \pi_m^i = \lambda_k h_{k,0} p_k^j. \quad (23)$$

Similar to the continuous-game scenario in Theorem 3, it is easy to verify that as $\lambda_k \rightarrow \infty$, $\pi_k^1 \rightarrow 1$ and $u_0 \rightarrow 0$. We note that $\forall \lambda$, the equation array (23) always has a solution since the mixed-strategy NE exists. With (18) and (23), we can obtain

$$u_0 = \sum_k \sum_{p_k^j \in \mathcal{P}_k} \frac{\psi(p_k^j, p-k) \pi_k^j \prod_{m \neq k, i} \pi_m^i}{h_{k,0} p_k^j}, \quad (24)$$

which must have a non-zero maximum value. Therefore, we can always find a finite λ_k that maximize (18) with the NE of

the follower game, Otherwise, it will contradict with the fact that $u_0 \rightarrow 0$ as $\lambda_k \rightarrow \infty$. \square

Following our discussion, we propose a pricing mechanism based on the myopic best response as Algorithm 2:

Algorithm 2 Heuristic price updating

Require: The MBS sets $\lambda_k = 0$ and the FUs arbitrarily initialize π_k^0 .

- 1: **while** the cross-tier SINR requirement (1) is not met **do**
- 2: **while** not converged **do**
- 3: $\forall k \in \mathcal{K}$, FU k updates π_k with (19)-(21).
- 4: **end while**
- 5: $\forall k \in \mathcal{K}$, FU k report π_k to the MBS.
- 6: The MBS announces the prices λ_k :

$$\lambda_k = \frac{\sum_{p \in \mathcal{P}} \psi_k(p_k, p-k) \prod_{i \in \mathcal{K}, j} \pi_i^j}{\sum_j h_{k,0} p_k^j \pi_k^j}, \quad (25)$$

- 7: **end while**
-

IV. SIMULATION RESULTS

The objective of this section is to provide insight into the impact of pricing on the network performance at the equilibrium, and the influence of strategy discretization on the learning process. In the simulation, we assume that the FAPs are randomly located indoors within a circle centered at the MBS with a radius of 300m. Each FU is placed within a circle centered at the corresponding FAP with a radius of 15m. The channel gains of the transmitter-receiver pairs are generated by a lognormal shadowing pathloss model with $h_{i,j} = d_{i,j}^{-k}$, in which k is the pathloss factor, $k = 4$ for the FUs and $k = 2.5$ for the MU. The parameters used in the simulation are summarized in Table I.

TABLE I
MAIN PARAMETERS USED IN THE FEMTOCELL NETWORK SIMULATION

Parameter	Value
Shared Bandwidth W	1MHz
Maximum MU transmit power p_0^{\max}	27dBm
Feasible region for FU transmit power $[p_k^{\min}, p_k^{\max}]$	$[0, 20]$ dBm
Additional FU circuit power p_a	3dBm
AWGN power $N_k, k = 0, \dots, K$	-40dBm
SINR threshold of the MU	3dB

A. Analysis of the Equilibrium in the Continuous Game

In the first simulation, we study the influence of the MBS price λ on the equilibrium of the follower subgame with 6 FAPs. For the convenience of demonstration, we suppose that the MBS charges an identical price to each FU-FAP link. Figure 1 provides the payoffs evolution of both the MBS and the FU-FAP links at the follower-game NE as the uniform price increases. We note in Figure 1.b that there exists an optimal value of λ to maximize the MBS revenue, which provides an experimental evidence of Theorem 3. We also note from Figure 1.a and 1.b that there exists a plateau region in which the average power efficiency remains almost the same

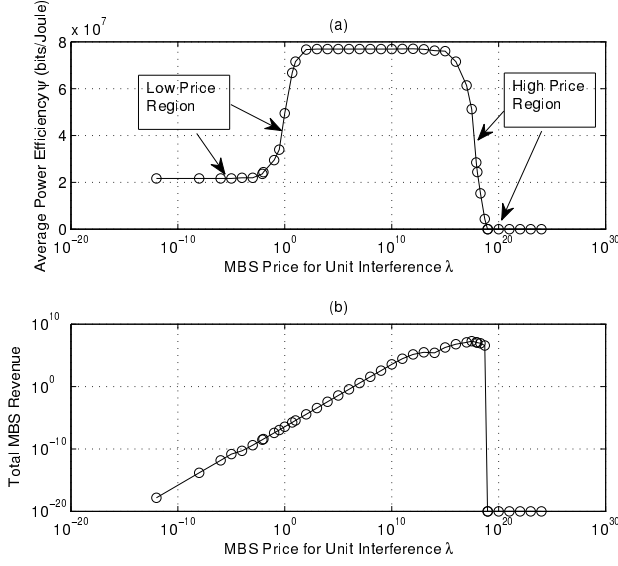


Fig. 1. Influence of the unit interference price on the NE of the continuous follower-game. (a) Average power efficiency at the NE vs. the unit price λ ; (b) total revenue collected by the MBS at the NE vs. the unit price λ .

while the MBS revenue keeps increasing. It means that without undermining the social welfare of the femtocells, the MBS is able to control the cross-tier interference by choosing the price within the plateau region.

We note from Figure 1 that at the optimal MBS price (i.e., the SE point), the FU performance dramatically degrades from the optimal condition. This leaves the room in practice for the MBS to trade a portion of the revenue for a socially better performance. Following the first simulation, we investigate the network performance at the proposed price (16) as the number of FAPs varies. In Figures 2 and 3, the user performance at the proposed price is compared to that at the accurate SE price, and that with no price ($\lambda = 0$). The accurate SE price is obtained using a semi-exhaustive searching method with bisection, and the utilities are obtained from Monte Carlo simulation. Figure 2 shows that at the proposed price a better FU performance can be achieved (Figure 2.b) at the cost of losing a significant portion of the MBS revenue (Figure 2.a). However, by measuring the expected SINR of the MU in Figure 3, we note that such trade-off is worthwhile since the performance deterioration of the MU is small when compared to the gain of the FU performance. Again, Figure 2.b and Figure 3 shows the fact that with no externality, the network performance can be heavily impaired (see curve “No price, $\lambda = 0$ ”).

B. Analysis of the Equilibrium in Discrete Game

Since we are interested in the impact of strategy discretization on the network performance, we adopt the same network setup in the first simulation for the discrete game. The parameter for the self-learning algorithm is given in Table II. To compensate for the divergence caused by the self-learning algorithm (Theorem 4), we examine the expected user utilities with Monte Carlo simulation. For the convenience of demonstration, we also suppose that the MBS places an uniform price. The simulation results are shown in Figure 4.

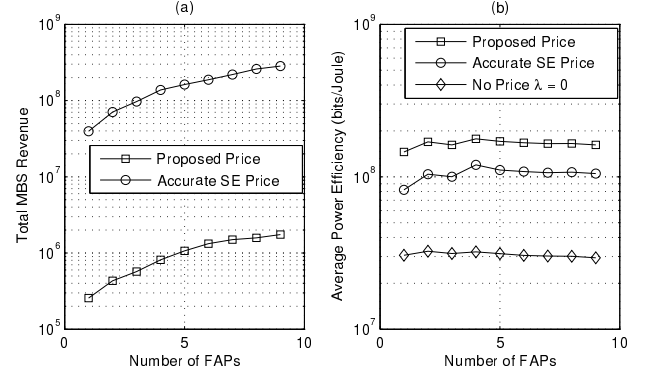


Fig. 2. MBS revenue and FU power efficiency at the SEs. (a) MBS revenue vs. the number of FAPs; (b) FU power efficiency vs. the number of FAPs.

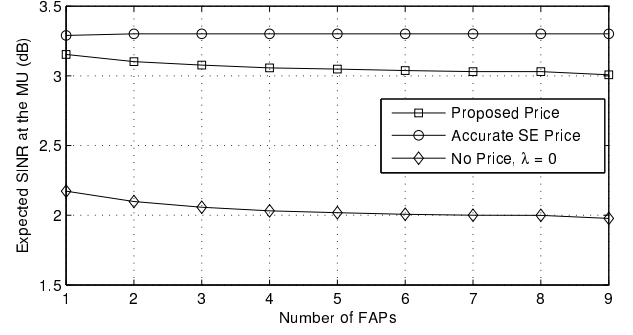


Fig. 3. Expected equilibrium SINR at the MU vs. the number of FAPs.

TABLE II
MAIN PARAMETERS USED IN THE SELF-LEARNING ALGORITHM

Parameter	Value
Boltzmann temperature τ	1
Learning rate for U_k^t and π_k^t	$1/t$ and $1/t^2$
Number of candidate power M	$M = 6$
Power sampling equation	$p_k^t = (1 - \frac{1}{M})p_k^{\min} + \frac{1}{M}p_k^{\max}$

Figure 4.b shows the existence of an optimal λ that maximizes the MBS revenue. It can be interpreted as an experimental evidence of Theorem 5. Comparing Figure 4.a and Figure 1.a, we note that in the discrete follower game (Figure 4.a), the “plateau” region extends to $\lambda \rightarrow 0$. It means that different from the case of continuous game (Figure 1.a), lacking an external price does not severely undermine the social performance of the FUs. However, the femtocell network may suffer from discretization of the power space and only achieve approximately 1/2 of the performance in the continuous game at most of the NEs.

Figure 5 shows the user performance at the proposed price with Algorithm 2, the accurate SE price and zero price, respectively. The comparison in Figure 5 shows that the FU performance at the proposed price is the best of the three. However, from Figure 5.b we note that without a pricing scheme, the FUs can still achieve a good performance level. As the number of FAPs increases, we can observe a deterioration in the FU performance. It means that mixed-strategies NE in the discrete game is not as good as the continuous-game NE in maintaining the performance when the size of network increases (see Figure 2.b).

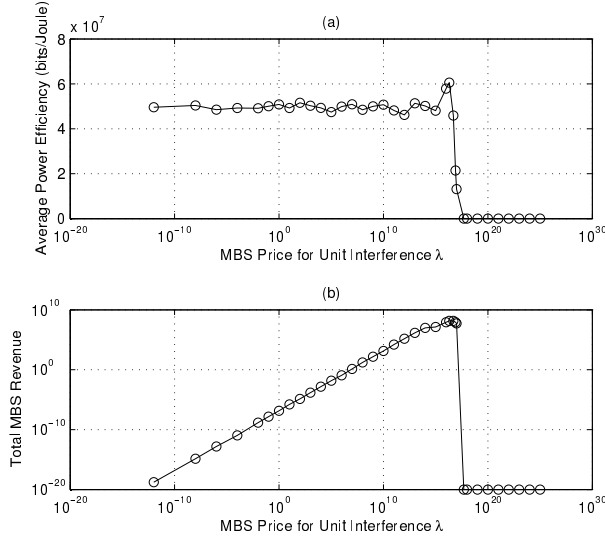


Fig. 4. Influence of the unit interference price on the NE of the discrete follower-game. (a) (Expected) average power efficiency at the NE vs. the unit price λ ; (b) total revenue collected by the MBS at the NE vs. the unit price λ .

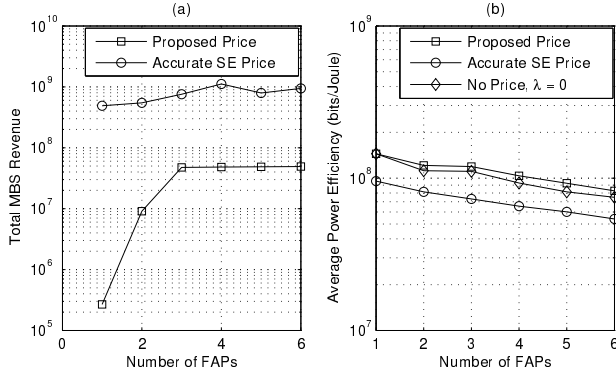


Fig. 5. Expected MBS revenue and FU power efficiency at the SEs in the discrete strategy space. (a) MBS revenue vs. the number of FAPs; (b) FU power efficiency vs. the number of FAPs.

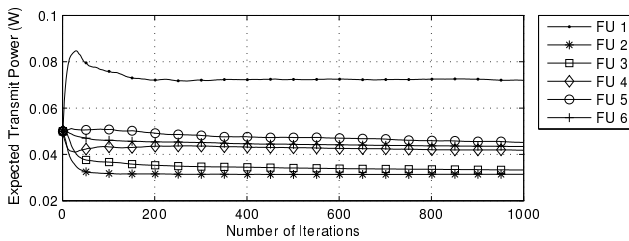


Fig. 6. Expected transmit power vs. the number of iterations.

Finally, we demonstrate in Figures 6 and 7 the convergence of the self-learning algorithm in the discrete game of 6 FUs. Figure 6 shows a snapshot of FUs' transmit-power evolution during the learning process of Algorithm 2. Figure 7 shows the corresponding strategy evolution of FU 1. By running the simulation for multiple times, we observe that most of the learning processes converge within 600 iterations.

V. CONCLUSION

In this paper, we have formulated the price-based power allocation problem in the two-tier femtocell network under the

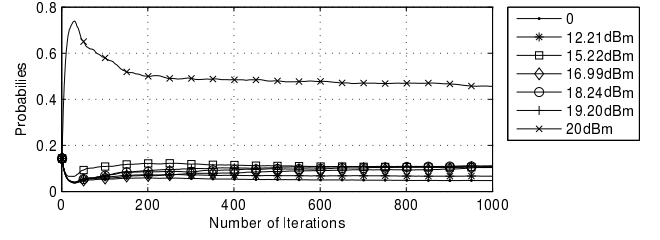


Fig. 7. Probability of power selection vs. the number of iterations by FU 1. framework of the Stackelberg game. We have provided the theoretical analysis of the properties of the equilibria in the scenarios of continuous FU power space and discrete power space, respectively. We have proposed two self-sufficient learning algorithms, one for each situation, for learning the NE of the follower game with limited information exchange. We have also provided the theoretical proof for the convergence of the learning algorithms. Our simulation results provides important insight in the different impact of the pricing mechanism on the network performance in the scenarios of the continuous game and the discrete game. In the simulation, we also show the efficiency of the proposed heuristic price-searching mechanism in the game. Our study provides an alternative way of designing the resource allocation protocols since no intercell information exchange is required in the femtocells.

APPENDIX A

PROOF OF THEOREM 1

Lemma 2 ([18]). *If a function $f(\mathbf{s})$ is twice differentiable, then supermodularity is equivalent to $\frac{\partial^2 f(\mathbf{s})}{\partial s_i \partial s_j} \geq 0 \forall s_i, s_j, j \neq i$.*

The first and the second conditions of a supermodular game in Definition 3 (trivially) holds for the proposed follower subgame (7). Then, Theorem 1 can be derived based on Lemma 2. By taking the component-wise derivative of $u_k(p_k, \mathbf{p}_{-k})$ with respect to p_k , we obtain:

$$\frac{\partial u_k}{\partial p_k} = -\frac{W \log(1+\gamma_k)}{(p_k + p_a)^2} + \frac{W G_k}{(1+\gamma_k)(p_k + p_a)} - \lambda_k h_{k,0}, \quad (26)$$

in which

$$G_k = \frac{h_{k,k}}{N_0 + h_{0,k}p_0 + \sum_{i \in \mathcal{K} \setminus \{k\}} h_{i,k}p_i}. \quad (27)$$

Then $\forall k, j \in \mathcal{K}, k \neq j$, the value of $\frac{\partial^2 u_k}{\partial p_k \partial p_j}$ is given by:

$$\frac{\partial^2 u_k}{\partial p_k \partial p_j} = \frac{W H_k p_k}{(1+\gamma_k)(p_k + p_a)^2} + \frac{W H_k \gamma_k}{(1+\gamma_k)^2(p_k + p_a)} - \frac{W H_k}{(1+\gamma_k)(p_k + p_a)}, \quad (28)$$

in which

$$H_k = \frac{h_{j,k} h_{k,k}}{\left(N_0 + h_{0,k}p_0 + \sum_{i \in \mathcal{K} \setminus \{k\}} h_{i,k}p_i\right)^2}. \quad (29)$$

It is easy to verify from (28) that $\frac{\partial^2 u_k}{\partial p_k \partial p_j} \geq 0$ if $\gamma_k \geq p_a/p_k$. By Lemma 2, u_k has increasing difference between p_k and any component of \mathbf{p}_{-k} if $\gamma_k \geq p_a/p_k$. By Definition 3, the proof of Theorem 1 is completed.

APPENDIX B PROOF OF LEMMA 1

The proof of Lemma 1 is derived from investigating the quasiconcavity of the FU payoff functions. For conciseness, the readers are referred to Sections 3.1 and 3.4 of [19] for the details of the definition in superlevel set and quasiconcavity.

Lemma 3 ([19]). *Suppose $f : \mathcal{D} \rightarrow \mathbb{R}$ be strictly quasiconcave where $\mathcal{D} \subset \mathbb{R}^N$ is convex. Then any local maximum of f on \mathcal{D} is also a global maximum of f on \mathcal{D} . Moreover, the set $\arg \max\{f(x) | x \in \mathcal{D}\}$ is either empty or a singleton.*

We first show that the utility function u_k in \mathcal{G}_f is quasiconcave. We examine the α -superlevel set of $u_k(p_k, p_{-k})$ in p_k , which is equivalent to the 0-superlevel set of $f_\alpha(p_k, p_{-k})$:

$$\mathcal{P}_{k,0} = \{p_k | f_\alpha(p_k, p_{-k}) \geq 0, 0 \leq p_k \leq p_k^{\max}, f_\alpha(p_k, p_{-k}) = W \log(1 + \gamma_k) - (\lambda_k h_{k,0} p_k + \alpha)(p_k + p_a)\}. \quad (30)$$

We note that $f_\alpha(p_k, p_{-k})$ is a concave function, so $\mathcal{P}_{k,\alpha}$ is convex by the definition of convexity. By the definition of quasiconcavity, $u_i(\lambda^*, p_k, p_{-k})$ is quasiconcave in p_k .

Then we show that $u_i(p_k, p_{-k})$ is strictly quasiconcave in p_k so the BR is a global maximum and thus single-valued. Without loss of generality, we consider the power allocation $\hat{p}_k \in [0, p_k^{\max}]$ with $u_k(\hat{p}_k, p_{-k}) = \alpha$. We assume that a different power allocation \tilde{p}_k satisfies $u_k(\tilde{p}_k, p_{-k}) \geq u_k(\hat{p}_k, p_{-k})$. Correspondingly, $f_\alpha(\hat{p}_k, p_{-k}) = 0$ and $f_\alpha(\tilde{p}_k, p_{-k}) \geq 0$. Observing the following condition for $f_\alpha(\hat{p}_k, p_{-k}) = 0$:

$$W \log(1 + G_k p_k) = (\lambda_s h_{k,0} p_k + \alpha)(p_k + p_a), \quad (31)$$

in which G_k is given in (27). We note that the right-hand side of (31) is a strictly increasing concave function and the left-hand side is a strictly increasing convex function in $[0, p_k^{\max}]$. Then the solution to $f_\alpha(\hat{p}_k, p_{-k}) = 0$ is unique in $[0, p_k^{\max}]$, so $f_\alpha(\tilde{p}_k, p_{-k}) > f_\alpha(\hat{p}_k, p_{-k})$. Based on the definition of concave function, the following inequality also holds for $0 < \delta < 1$:

$$\begin{aligned} f_\alpha(\delta \hat{p}_k + (1 - \delta) \tilde{p}_k) &\geq \delta f_\alpha(\hat{p}_k) + (1 - \delta) f_\alpha(\tilde{p}_k) \\ &> f_\alpha(\hat{p}_k) = 0. \end{aligned} \quad (32)$$

Therefore, the condition for strict quasiconcavity holds as $u_k(\delta \hat{p}_k + (1 - \delta) \tilde{p}_k) > \min(u_k(\tilde{p}_k, p_{-k}), u_k(\hat{p}_k, p_{-k})) = \alpha$. Then Lemma 1 is a direct conclusion based on Lemma 3.

APPENDIX C PROOF OF THEOREM 2

Lemma 4 ([7]). *If the best-response functions of a non-cooperative game \mathcal{G} are standard functions for all the players, then the game has a unique NE in pure strategies.*

Observing (4), we note that the maximum net-payoff function is lower-bounded by 0 with $p_k = 0$. Since the power vector is always nonnegative, the property of positivity in the BR for each FU k immediately follows Lemma 1.

We denote $I_k(p_{-k}) = (N_k + h_{0,k} p_0 + \sum_{j \in \mathcal{K} \setminus \{k\}} h_{j,k} p_j) / h_{k,k}$. Noting that $I_k(p_{-k})$ is a strictly increasing function of p_{-k} , monotonicity of $\hat{p}_k(p_{-k})$ can be illustrated by proving that

function $p_k(I_k)$ is monotonically increasing in I_k . From (26) we obtain the necessary condition for p_k to be the BR as $\frac{\partial u_k}{\partial p_k} = 0$ which is equivalent to

$$\begin{aligned} \omega(p_k, I_k) &= \frac{W(p_k + p_a)}{I_k} - W(1 + \frac{p_k}{I_k}) \log(1 + \frac{p_k}{I_k}) \\ &\quad - \lambda_k h_{k,k} (p_k + p_a)^2 (1 + \frac{p_k}{I_k}) = 0. \end{aligned} \quad (33)$$

Since $\frac{\partial p_k}{\partial I_k} = -\frac{\partial \omega}{\partial I_k} / \frac{\partial \omega}{\partial p_k}$, we have

$$\frac{\partial \omega}{\partial I_k} = \frac{1}{I_k^2} \left(\xi(p_k) + W p_k \log(1 + \frac{p_k}{I_k}) - W p_a \right), \quad (34)$$

in which $\xi(p_k) = \lambda_k h_{k,k} (p_a^2 p_k + 2 p_a p_k^2 + p_k^3)$, and

$$\frac{\partial \omega}{\partial p_k} = -\frac{1}{I_k} \left(\zeta(p_k) + W \log(1 + \frac{p_k}{I_k}) \right) \quad (35)$$

in which $\zeta(p_k) = \lambda_k h_{k,k} (p_a + p_k)(p_a + 2I_k + 3p_k)$. We note that $\frac{\partial \omega}{\partial p_k} < 0$, then the property of monotonicity holds iff $\frac{\partial \omega}{\partial I_k} \geq 0$. With the inequality of logarithmic function [23], $\log(1 + x) \geq x/(1 + x)$ for $x \geq -1$, we obtain

$$\frac{\partial \omega}{\partial I_k} \geq \frac{1}{I_k^2} \left(\xi(p_k) + \frac{W}{I_k + p_k} (p_k^2 - p_a(I_k + p_k)) \right). \quad (36)$$

Therefore, $\frac{\partial p_k}{\partial I_k} \geq 0$ if $p_k^2 - p_a(I_k + p_k) \geq 0$. Then we obtain the condition for $\hat{p}_k(p_{-k})$ to be monotonic as:

$$\frac{h_{k,k} p_k}{N_k + h_{0,k} p_0 + \sum_{j \in \mathcal{K}} h_{j,k} p_j} \geq \frac{p_a}{p_k}. \quad (37)$$

The proof of scalability is based on Lemma 1. According to Lemma 1, there is a one-to-one correspondence between \hat{p}_k and $\hat{\gamma}_k$. We define $J_k(p_{-k}) = \sum_{j \in \mathcal{K} \setminus \{k\}} h_{j,k} p_j$, then from (2) the BR can be written as

$$\hat{p}_k(p_{-k}) = \frac{\hat{\gamma}_k (N_0 + h_{0,k} p_0 + J_k(p_{-k}))}{h_{k,k}}. \quad (38)$$

From $\frac{\partial u_k}{\partial p_k} = 0$, we can prove that $\frac{\partial \gamma_k}{\partial J_k} \leq 0$ with the same technique as proving monotonicity (which is omitted for conciseness). Therefore, γ_k is a decreasing function of J_k . Since $J_k(p_{-k})$ is a standard function [18], we realize that if $\alpha > 1$, $\hat{\gamma}_k(\alpha p_{-k}) \leq \hat{\gamma}_k(p_{-k})$ and $J_k(\alpha p_{-k}) \leq \alpha J_k(p_{-k})$. Then, monotonicity holds for $\hat{p}_k(p_{-k})$ since

$$\hat{p}_k(\alpha p_{-k}) \leq \frac{\hat{\gamma}_k(p_{-k}) (N_0 + h_{0,k} p_0 + \alpha J_k(p_{-k}))}{h_{k,k}} \leq \alpha \hat{p}_k(p_{-k}). \quad (39)$$

Therefore, $\hat{p}_k(p_{-k})$ is a standard function. Based on Lemma 4, the NE of the follower subgame (7) is unique.

APPENDIX D

THE DERIVATION OF ASYMPTOTIC BEHAVIOR MODELS

The necessary condition for the NE of the FUs is given by (13). As $\lambda_k \rightarrow 0$, the solution of the BRs in the follower game will be independent of λ_k and can be approximated by

$$(1 + \gamma_k) \log(1 + \gamma_k) - W \gamma_k - W G_k p_a = 0, k \in \mathcal{K}, \quad (40)$$

in which G_k is given by (27). From Lemma 1, the solution to (40) is unique. Then the payment by FU k will be a linear function of λ_k , $r_k = h_{k,0}\hat{p}_k\lambda_k$.

As $\lambda_k \rightarrow \infty$, $\forall k \in \mathcal{K}$, $p_k \rightarrow 0$. From (13) we obtain:

$$h_{k,k}\lambda_k(p_k + p_a) = \frac{Wh_{k,k}}{I_k + h_{k,k}p_k} - \frac{W \log(1 + \frac{h_{k,k}p_k}{I_k})}{(p_k + p_a)}, \quad (41)$$

in which I_k is the sum of interference plus noise defined in Appendix C. With $p_k \rightarrow 0$, $\forall k$, (41) can be approximated by:

$$h_{k,k}\lambda_k(p_k + p_a) \approx \frac{Wh_{k,k}}{N_k + h_{0,k}p_0}. \quad (42)$$

From (42) we obtain

$$p_k \approx \frac{W}{\lambda_k(N_k + h_{0,k}p_0)} - p_a. \quad (43)$$

Based on (40) and (43) we obtain the asymptotic models (14) and (15) for the FU behaviors.

APPENDIX E

PROOF OF THEOREM 4

Lemma 5 ([21]). *Consider game \mathcal{G} with payoff function $u_k(\mathbf{s})$ for player k . If the sequence of stochastic fictitious play converges, it also holds for game $\tilde{\mathcal{G}}$ with payoff function $\tilde{u}_k(\mathbf{s}) = \kappa_k u_k(\mathbf{s}) + \vartheta(s_k)$, in which κ_k is a positive constant and $\vartheta(s_k)$ only depends on player k 's own behavior.*

Lemma 6 ([21]). *Consider stochastic fictitious play starting from any arbitrary π^0 . If \mathcal{G} is a supermodular game, then*

$$\Pr(\omega\{\pi_k^0\} \subset RP \text{ or } \omega\{\pi_k^0\} \subset M_i \cap [\underline{\pi}_k, \bar{\pi}_k] \text{ for } k) = 1$$

where $\omega\{\pi_k^0\}$ is an invariant set of the solution trajectory starting from π_k^0 . RP is the set of rest points (fixed points) and $\underline{\pi}_k, \bar{\pi}_k \in RP$ such that $RP \subset [\underline{\pi}_k, \bar{\pi}_k]$. M_i is a finite Lipschitz submanifold and every persistent non-convergence trajectory is asymptotic to one in M_i .

The proof of Theorem 4 starts by investigating the property of supermodularity in the FU subgame. With discrete power set, the expected payoff function (17) can be rewritten as:

$$u_k(\pi_k, \pi_{-k}, \lambda) = \sum_{\mathbf{p} \in \mathcal{P}} \psi_k(p_k, p_{-k}) \prod_{i \in \mathcal{K}, j} \pi_i^j - \sum_j \lambda_k h_{k,0} p_k^j \pi_k^j, \quad (44)$$

which is in the form $u_k(\pi) = \kappa_k \tilde{u}_k(\pi) + \vartheta(\pi_k)$ with $\kappa_k = 1$. By Lemma 5, we only need to check the game with payoff

$$\tilde{u}_k(\pi_k, \pi_{-k}, \lambda) = \sum_{\mathbf{p} \in \mathcal{P}} \psi_k(p_k, p_{-k}) \prod_{i \in \mathcal{K}, j} \pi_i^j. \quad (45)$$

With Definition 3 and Lemma 2, it is easy to check that the game with payoff function (45) and mixed strategies is a supermodular game. Based on Theorem 6 of [20], the process of π_k^t produced by (19)-(21) will almost surely be an asymptotic pseudotrajectory of the smooth BR dynamics:

$$\dot{\pi}_k = \beta_k(\pi_{-k}) - \pi_k.$$

Then, based on Lemma 6, for the game with payoff (45) stochastic fictitious play almost surely converges with any arbitrary initial π^0 . With Lemma 5, the convergence holds for the original subgame with payoff u_k , so Theorem 4 is proved.

REFERENCES

- [1] J. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas in Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012.
- [2] C. Saraydar, N. B. Mandayam, and D. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Trans. on Commun.*, vol. 50, no. 2, pp. 291–303, Feb. 2002.
- [3] A. MacKenzie and S. Wicker, "Game theory in communications: motivation, explanation, and application to power control," in *Proc. IEEE GLOBECOM*, vol. 2, San Antonio, TX, Dec. 2001, pp. 821–826.
- [4] N. Feng, S.-C. Mau, and N. B. Mandayam, "Pricing and power control for joint network-centric and user-centric radio resource management," *IEEE Trans. on Commun.*, vol. 52, no. 9, pp. 1547–1557, Sep. 2004.
- [5] H. Yu, L. Gao, Z. Li, X. Wang, and E. Hossain, "Pricing for uplink power control in cognitive radio networks," *IEEE Trans. Veh. Tech.*, vol. 59, no. 4, pp. 1769–1778, May 2010.
- [6] J. Huang, R. Berry, and M. Honig, "A game theoretic analysis of distributed power control for spread spectrum ad hoc networks," in *Proc. Inter. Sym. Info. Thm.*, Adelaide, Australia, Sep. 2005, pp. 685–689.
- [7] Z. Han, D. Niyato, W. Saad, T. Basar, and A. Hjørungnes, *Game theory in wireless and communication networks*. Cambridge University Press, 2012.
- [8] J. Zhang and Q. Zhang, "Stackelberg game for utility-based cooperative cognitiveradio networks," in *Proc. ACM MobiHoc*, New York, NY, May 2009, pp. 23–32.
- [9] X. Kang, R. Zhang, and M. Motani, "Price-based resource allocation for spectrum-sharing femtocell networks: A stackelberg game approach," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 538–549, Apr. 2012.
- [10] Y. Wu, T. Zhang, and D. H. K. Tsang, "Joint pricing and power allocation for dynamic spectrum access networks with stackelberg game model," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 12–19, Jan. 2011.
- [11] L. Yang, H. Kim, J. Zhang, M. Chiang, and C.-W. Tan, "Pricing-based spectrum access control in cognitive radio networks with random access," in *Proc. IEEE INFOCOM*, Shanghai, China, Apr. 2011, pp. 2228–2236.
- [12] P. Zhou, Y. Chang, and J. Copeland, "Reinforcement learning for repeated power control game in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 54–69, Jan. 2012.
- [13] A. Galindo-Serrano, L. Giupponi, and M. Dohler, "Cognition and docation in ofdma-based femtocell networks," in *Proc. IEEE GLOBECOM*, Miami, FL, Dec. 2010, pp. 1–6.
- [14] C. Long, Q. Zhang, B. Li, H. Yang, and X. Guan, "Non-cooperative power control for wireless ad hoc networks with repeated games," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, pp. 1101–1112, Aug. 2007.
- [15] M. Bennis, S. Perlaza, P. Blasco, Z. Han, and H. Poor, "Self-organization in small cell networks: A reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3202–3212, Jul. 2013.
- [16] X. Chen, H. Zhang, T. Chen, and M. Lasanen, "Improving energy efficiency in green femtocell networks: A hierarchical reinforcement learning framework," in *Proc. IEEE ICC*, Budapest, Hungary, Jun. 2013, pp. 2241–2245.
- [17] C. Han, T. Harrold, S. Armour, I. Krikidis, S. Videv, P. M. Grant, H. Haas, J. Thompson, I. Ku, C. Wang, T. A. Le, M. Nakhai, J. Zhang, and L. Hanzo, "Green radio: radio techniques to enable energy-efficient wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 46–54, Jun. 2011.
- [18] E. Altman and Z. Altman, "S-modular games and power control in wireless networks," *IEEE Trans. Auto. Ctrl.*, vol. 48, no. 5, pp. 839–842, May 2003.
- [19] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [20] D. S. Leslie and E. Collins, "Convergent multiple-timescales reinforcement learning algorithms in normal form games," *Ann. Appl. Probab.*, vol. 13, no. 4, pp. 1231–1251, Feb. 2003.
- [21] J. Hofbauer and W. H. Sandholm, "On the global convergence of stochastic fictitious play," *Econometrica*, vol. 70, no. 6, pp. 2265–2294, Nov. 2002.
- [22] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.
- [23] F. Topsok, "Some bounds for the logarithmic function," in *Inequality theory and applications*, vol. 7, no. 2. Nova Science Pub Incorporated, 2006, pp. 137–156.